Validation of reference forecasts for passenger transport

Matts Andersson, Karin Brundell-Freij, Jonas Eliasson (*presenter*)

Summary
This paper compares a number of forecasts for aggregate passenger transport (by mode) produced in Sweden between 1975 and 2009 with actual outcomes. In addition to the comparison between forecasts and outcomes, we explore to what extent forecast errors are due to erroneous input assumptions, enabling us to give a more fair judgment of the models' forecasting abilities.

We find substantial differences between forecasts and actual outcomes. In forecasts produced since the early 1990s, road and air traffic growth rates have generally been overpredicted. Aggregate railway growth has been fairly accurate, but commercial long-distance railway growth has been overpredicted, and the growth of subsidized intra-regional railway travel has been underpredicted (following vast unanticipated supply increases).

When discussing forecasting errors, it is useful to distinguish between different sources of errors:

- Model deficiencies: deficiencies in underlying theories, estimation methodologies, model implementation, data used for estimation and so on.
- Differences between cross-sectional and intertemporal relationships: most transport models that are able to generate detailed forecasts are estimated based on cross-sectional data. Using such models to predict future reference forecasts thus rests on the tacit assumption that cross-sectional and intertemporal relationships are equal.
- Changes in preferences or behavior: using models to predict the future also tacitly assumes stability in preferences and behavior.
- Assumption errors: to produce a reference forecast, it is necessary to make a large number of assumptions about future transport supply and general societal and socioeconomic variables.

Focusing on car traffic forecasts, we show that a very large share of forecast errors can be explained by assumption errors, i.e. input variables turning out to be different than what was assumed when the forecasts were made. Even the original forecasts are much closer to actual outcomes than simple trendlines would have been, and once the input assumptions are corrected, the forecasts vastly outperform simple trendlines.

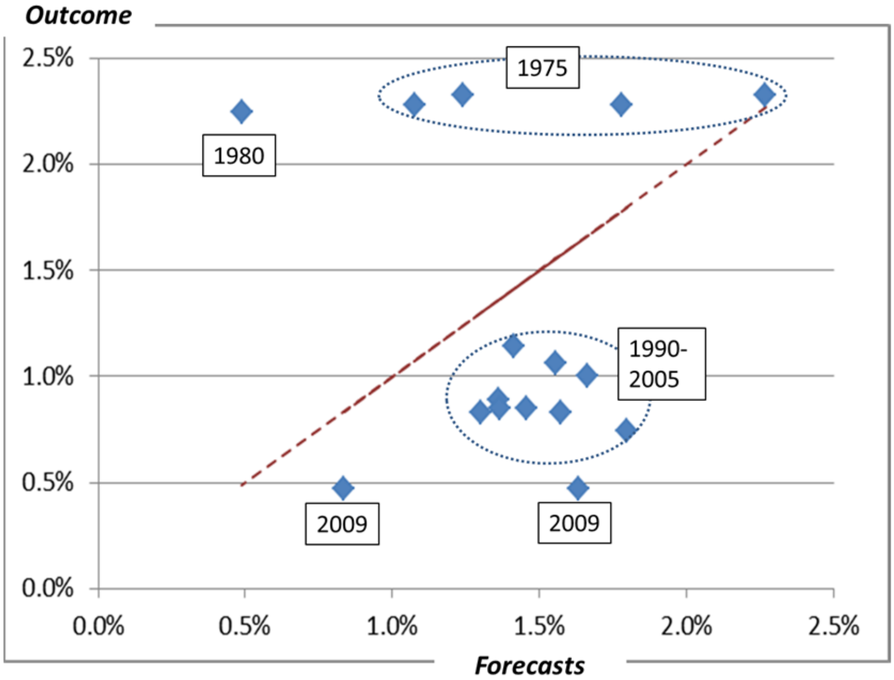
The models used to produce the forecasts from the last decades are state-of-the-art (at the time) nested logit models, estimated on disaggregate cross-sectional data, and producing detailed forecasts on network and origin-destination levels. The present paper focuses on how well the models have managed to predict aggregate numbers of total passenger kilometers by mode some time into the future (typically more than a decade). The fact that they vastly outperform simple time-series trendline forecasts of aggregate passenger transport is interesting: it indicates that the potential problems of using cross-sectional models for forecasting intertemporal changes seem to be limited. This tentative conclusion is also supported by the finding that elasticities from the cross-sectional models are consistent with those from a time-series model.

Background
Transport forecasts play an important role in policy design, decision-making, and public debate. The question addressed in this paper is how accurate forecasts for passenger transport are and to what extent forecast errors can be explained by wrong assumptions about input variables. When discussing the validation of transport forecasts, it is useful to distinguish between

reference forecasts, which predict transport volumes in a future year in a baseline scenario, and *policy* forecasts, which predict the effects of some policy by comparing "do-nothing" and "do-something" scenarios. Such "policies" can be any intervention in the transport system, including infrastructure investments, changes in prices or taxes, new regulations, and so on. In practice, a *project* forecast – such as the forecasted traffic volume on a planned road – is a mixture of the two in which a reference forecast gives the baseline traffic volumes in the do-nothing scenario (which is usually several years into the future), and a policy forecast predicts how traffic volumes will be changed by the new road. When analyzing forecast accuracy, however, it is useful to distinguish between the two for several reasons. In this paper, we analyze reference forecasts.

Results

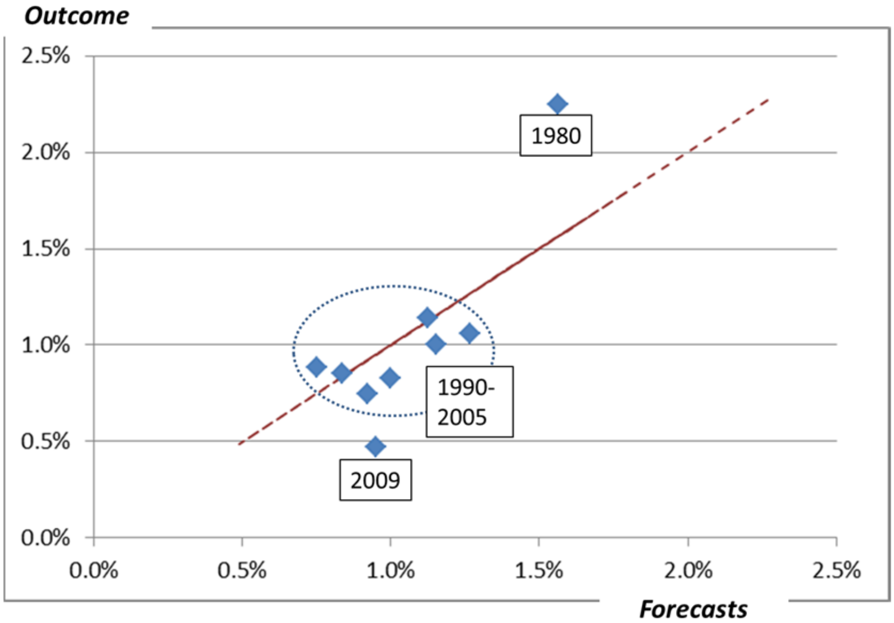
The following diagram summarizes some of the main comparisons between forecasts and outcomes for aggregate private car transport. The scattergram shows predicted yearly traffic growth (x-axis) compared to actual yearly traffic growth (y-axis). Numbers in boxes show the year the forecast was produced. Forecast periods are typically a decade or two into the future.



All forecasts from 1990 have overpredicted car traffic growth, on average by a factor 2. Predicted growth factors are clustered around 1.5% while actual growth factors are clustered around 0.7–0.8%. There is no apparent time trend in the forecasts, and the overpredictions have the same magnitude for all forecasts. (Note that most forecasts have been produced with different models.)

(Similar comparisons for railway transport and local public transport are omitted from this extended abstract for space reasons.)

To explore to what extent forecasts errors are due to input assumption errors, we adjust the original forecasts with respect to assumption errors in fuel price, fuel economy, car ownership, GDP/capita and total population. After doing this (for those forecasts where input assumptions were documented), we get the following comparison between (adjusted) forecasts and actual outcomes:



In conclusion, a very large share of total forecast errors can be explained by input variables being different in reality from what was assumed in the forecasts. This, together with the fact that even the original forecasts perform much better than simple trendlines (and the adjusted forecasts much more so), is a testimony to the explanatory power of the models. This result, and the consistency between elasticities from time series and cross-sectional models (omitted from this extended abstract for space reasons), indicate that the potential problems of using cross-sectional models for forecasting intertemporal changes are in fact limited.

Conclusions

Reference forecasts form an important baseline for analyzing individual policies, in particular designing and evaluating infrastructure investments. Such analyses ideally need a best estimate of the most likely future situation. The conclusion that most of the observed forecasting errors seem to be explained by assumption errors is comforting in the sense that models are broadly trustworthy – but it does not in itself make the choice of input variables easier. While we can draw some conclusions on how to improve the choice of what input variables to assume, our results highlight the need for extensive and systematic sensitivity analyses with respect to input variables.